

World Wide Web-Based Access to Heterogenous Information Resources for Cytokine Research and Education

Gary K. Allen, D.V.M., Ph.D.,^{1,2} Timothy B. Patrick, Ph.D.,^{1,3} Michael P. Murtaugh, Ph.D.⁴

¹Medical Informatics Group, School of Medicine,

²Dept. of Veterinary Pathobiology, College of Veterinary Medicine, and

³School of Library and Information Sciences, University of Missouri, Columbia, MO 65211 and

⁴Dept. of Veterinary Pathobiology, University of Minnesota, St. Paul, MN 55108

Cytokines are currently the focus of intense research efforts, and information related to cytokine biology is expanding at an extremely rapid rate. The primary literature has become insufficient as a source for timely information, and efficient searches of the older literature are complicated by the nature of cytokine nomenclature. A few computer-accessible cytokine-related data sources have recently been developed, but these are typically limited in scope and not integrated with other information resources. We have developed Cytokine Explorer, a World Wide Web (WWW) - based resource to provide access to and guidance in use of information for research and education in cytokine biology. The resource consists of: (1) a searchable thesaurus of cytokine nomenclature; (2) a cytokine reagent database containing information from academic institutions and commercial sources; (3) a set of links to cytokine-related WWW sites; and (4) a database of functional information retrieval maps for guided retrieval of cytokine information from combinations of heterogenous sources.

INTRODUCTION

Timely, integrated access to information from heterogenous sources has become important for biomedical research and education. In the field of cytokine biology, expanded research efforts have dramatically increased the generation of information, which has resulted in a rapid expansion of the scientific literature. Currently available computerized tools for cytokine information retrieval are limited with respect to user interface, information types available, and information integration capabilities.

The World Wide Web (WWW) has emerged as a powerful information delivery platform, and WWW applications have been developed for clinical and biomedical research information needs.¹⁻⁴ We hypothesized that cytokine research and education could be enhanced by a WWW-based resource that provided an integrated, user-friendly method for information acquisition from sources across the Internet. In this paper, we report the development of *Cytokine Explorer*, a WWW-based resource designed to provide timely, efficient, and integrated access to disparate cytokine information sources.

BACKGROUND

Cytokines are a large, diverse group of soluble cell-derived proteins and peptides that act as biological mediators at extremely low (nano- to picomolar) concentrations and modify both normal and pathological activities of individual cells and tissues. Almost all cytokines mediate multiple biological activities and different cytokines frequently manifest overlapping biological activity. Cytokines play a central functional role in virtually all host responses, including the regulation of immune responses to infectious and neoplastic diseases, inflammation, hematopoiesis, wound healing, embryo- and organogenesis, and neuroimmunological and neuroendocrinological regulation. Elucidation of biological activities of individual cytokines has prompted clinical investigations targeted at development of novel therapeutic and preventive interventions, particularly in the areas of infectious disease, hematopoietic dysfunction, and tumor therapy. The burgeoning understanding of the mechanisms governing cytokine actions thus represents an important contribution to biomedical knowledge.⁵

The large number of cytokines, their multiple, often overlapping activities, and the explosive rate of expansion of the primary cytokine literature have created an environment where maintenance of intellectual currency is extremely difficult, even for experts of cytokine biology. Factors that are particularly relevant to retrieval attempts in this domain include (1) the nature of cytokine nomenclature, (2) the requirement for timely information, and (3) the current lack of a method for integrated access to heterogenous information sources.

The problem of cytokine nomenclature

Retrieval of cytokine-related information is complicated by cytokine nomenclature, which a cytokine lexicographer has recently characterized as "a zoo of factors in a jungle of interactions surrounded by deep morasses of acronyms and bleak deserts of synonyms."⁶ Most cytokine names were originally derived from their observed *in vitro* activities, and then often reduced to acronyms (e.g., T cell growth

factor [TCGF]). This practice was partly responsible for the accumulation of synonyms for some cytokines. For example, there are more than 40 recognized synonyms for interleukin-1. This functional classification became confounded upon recognition that several discrete, uniquely-named soluble factors could mediate identical activities. Further complications have occurred as subsequent research has revealed that many of the original cytokine names are actually misnomers, and do not reflect the factor's authentic biological activity.⁶

Current attempts at consolidating the cytokine nomenclature focus on use of an enumerated scheme. For example, interleukins are currently numbered up to 17 (IL-17). This system has yet to be consistently applied, however, and names based on functional activities still predominate for numerous cytokine subsets.⁶ Thus, while enumeration may eventually result in enhanced precision in cytokine nomenclature, systematic searching of the primary cytokine literature demands expert knowledge of the biological characteristics of cytokines, as well as an awareness of both current and historic cytokine nomenclature.

The problem of timely information retrieval for cytokines

The primary literature is the most widely-used source of information on cytokines, but the sheer volume and lack of currency of this media makes it unsuitable for individuals needing high-quality, timely information. The rate of growth of the cytokine literature is reflected by the current (March, 1996) Medical Subject Headings (MeSH)⁷ for cytokines, which include headings for interleukins 1-13. The discovery of interleukins 14 through 17 has yet to be reflected in MeSH. In response to demand for timely information, some publishers provide electronic access to the primary immunology literature.⁸⁻¹⁰ While some of these offer full-text versions of articles, most provide only titles or abstracts.

Compilations and reviews of the scientific literature on cytokine biology are regularly published, but the utility of these reviews is diminished by the extremely rapid rate of advancement in this field. Thus, investigators are increasingly dependent upon research meetings and continuous, comprehensive survey of the primary literature for access to timely information.

Recently, some efforts have been directed toward the development of current computerized databases for cytokine biology.¹¹⁻¹⁴ Although timely information may be retrieved from several WWW sites, most of these sites provide information on some discrete aspect of cytokine biology. For example, there are

sites providing information such as 3-dimensional structures of cytokines and cytokine receptors,¹² or access to a cytokine cDNA database.¹⁴ Thus, available cytokine-related sites tend to serve special interests by offering only limited types of information on cytokines.

The problem of access to heterogenous information sources

Cytokine information retrieval can be complex because it frequently requires access to a combination of multiple, often heterogenous, information sources. For example, a researcher who has isolated a cell-derived factor for which the amino acid sequence and gene nucleic acid sequence have been determined could need to ascertain whether this factor is indeed a novel cytokine. This researcher would logically require access to protein and nucleic acid sequence databases, as well as the primary cytokine literature (Figure 1).

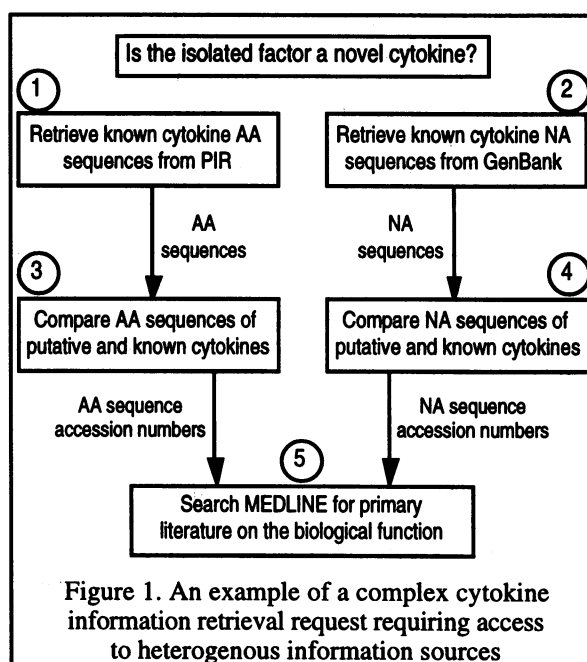


Figure 1. An example of a complex cytokine information retrieval request requiring access to heterogenous information sources

Importantly, answering this research question would require integrated access, through data connections, to all of these information sources. A data connection exists between two information sources when data contained in the output of one source may, under a transformation, be used as input to the other.¹⁵ Thus, success may be realized only upon iterative and/or linked searches wherein the result from one search (*e.g.*, an accession number from a sequence database) is used as a search parameter in another (*e.g.*, for primary literature in MEDLINE).

In the given example, sequence information is used to link the sequence databases to MEDLINE. In step 1, amino acid (AA) sequences of known cytokines are retrieved from the Protein Identification Resource

(PIR) database.³ In step 2, the nucleic acid (NA) sequences of known cytokines are retrieved from the NCBI Genbank database.⁴ In steps 3 and 4, the AA and NA sequences of the newly isolated factor and those of known cytokines are compared for structural similarity. The data links between step 1 and step 3 and between step 2 and step 4 are AA sequence and NA sequence, respectively. In step 5, MEDLINE is searched for primary literature about selected biological functions of the cytokines. The data links between steps 3 and 5 and between steps 4 and 5 are AA sequence accession number and NA sequence accession number, respectively.

Knowledge engineering is required to develop an informatics representation that will facilitate a complex retrieval in a given information domain.¹⁵ The representation incorporates expert knowledge of the domain, in the form of relevant retrieval questions, as well as knowledge of the available and appropriate information sources and how to combine them. Figure 1 is an example of an information retrieval map, a representation that may form the basis for a graphical user interface.

A system which addressed these three problems of access to cytokine information resources would be of great benefit to cytokine research and education. We are developing Cytokine Explorer,¹⁶ an integrated, user-friendly cytokine information system, to address these problems.

SYSTEM DESCRIPTION

Design

We first defined the types of cytokine information required and identified sources from which that information could be obtained. We focused on identifying sources of information for the preliminary datatypes listed in Table 1. Comparable datatypes were identified for other cytokine-related research reagents, including anti-cytokine antibodies and nucleic acid reagents (e.g. PCR primers, hybridization probes). We used standard literature searches, Internet searches with commercially-available search engines, and personal contacts to identify individual, institutional, and commercial WWW-based cytokine information sources.

Next, we characterized and functionally organized the resources identified. This included construction of:

- (1) a searchable cytokine reagent database that uses multiple pick-lists (cytokine name, species, and reagent type) for user input;
- (2) a searchable thesaurus of cytokine nomenclature, which uses free text input;
- (3) a set of information retrieval maps to support complex retrieval of cytokine information from WWW-based resources;

(4) links to other WWW-accessible cytokine information sources, including a direct link to the University of Minnesota veterinary cytokine database for data submission.

(5) a user evaluation form, to obtain input from end-users regarding attitudes toward services provided and to elicit suggestions for system enhancements.

Table 1. Relevant datatypes for cytokines

| Source Species | Function |
|--------------------------------|---------------------------------|
| <u>Classification</u> (family) | Source cells |
| <u>Nomenclature</u> | Agonists |
| <u>Genetics</u> | Antagonists |
| Nucleic acid sequence | Target cells |
| Chromosomal location | Species |
| <u>Structure</u> | Cell type |
| Size | Biological effects |
| No. of amino acids | Major |
| Molecular weight | Minor |
| Theoretical | <u>References</u> |
| Observed | <u>Sources and Availability</u> |
| Amino acid sequence | |
| Glycosylation state | |
| Quaternary structure | |

We use ODBC to access both the reagent database and the thesaurus (Figure 2). The information retrieval maps are implemented in two ways: (1) by using GIF files as client-side image maps; and (2) by creating the maps with a Java applet.

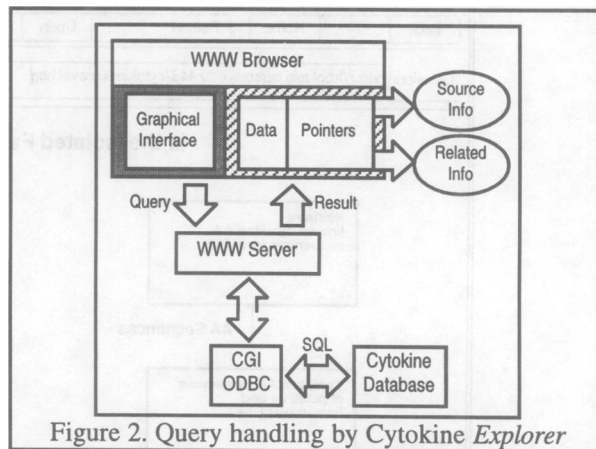


Figure 2. Query handling by Cytokine Explorer

Performance

Figure 3 illustrates sample output generated from a request to the cytokine reagent database of Cytokine Explorer. Results are formatted according to the specific query type and returned to the user's WWW browser. Results include the requested reagent data, relevant literature citations, and the name, address, phone and FAX numbers of the reagent source. Direct links to the reagent source (e.g., email link for individuals, URL's for corporate materials) are also

provided. Dynamically provided links permit retrieval of additional, related information from other information sources. For example, nucleic acid results include a link to the NCBI GenBank database. This link automatically includes the species and cytokine name, selected by the user in the *Explorer* interface, as parameters searching the NCBI GenBank database. Using similar forms, the system responds to queries to the cytokine nomenclature thesaurus with listings of currently accepted cytokine synonyms. Nomenclature data is based upon a recently published cytokine dictionary.⁶

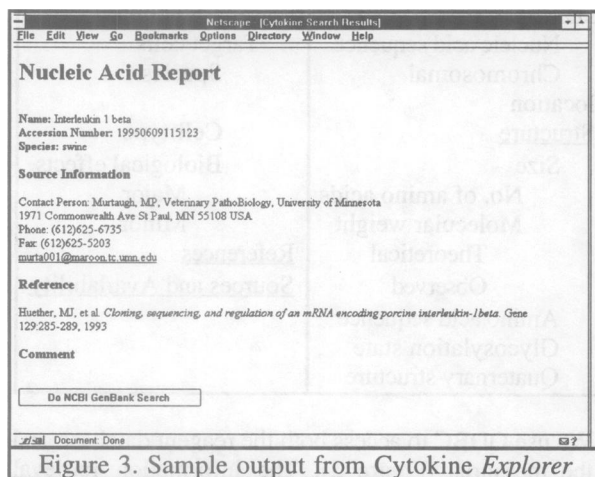


Figure 3. Sample output from Cytokine Explorer

The information retrieval maps provide graphic encapsulation of common complex retrieval requests for cytokine information. Currently, data are moved through the multiple nodes of the map by cutting and pasting. Figure 4 depicts a functional map created by a Java applet. Each node is associated with a URL. For example, the upper right node is associated with the URL for NCBI's GenBank service.

DISCUSSION

The lack of appropriate electronic information sources has complicated retrieval of cytokine information, and made it impossible to accomplish with any single currently available information retrieval system. In order to achieve effective, efficient information retrieval in this field, multi-domain expertise and knowledge must be used to access multiple information sources. Requisite information sources frequently differ with regard to network location, syntax, operating system, and hardware platforms. Distributed computing holds great potential for solving the problems associated with cytokine information retrieval. Specialized and optimized distributed computing platforms that work in concert with each other are becoming increasingly necessary to meet the complex informational needs of cytokine biologists.

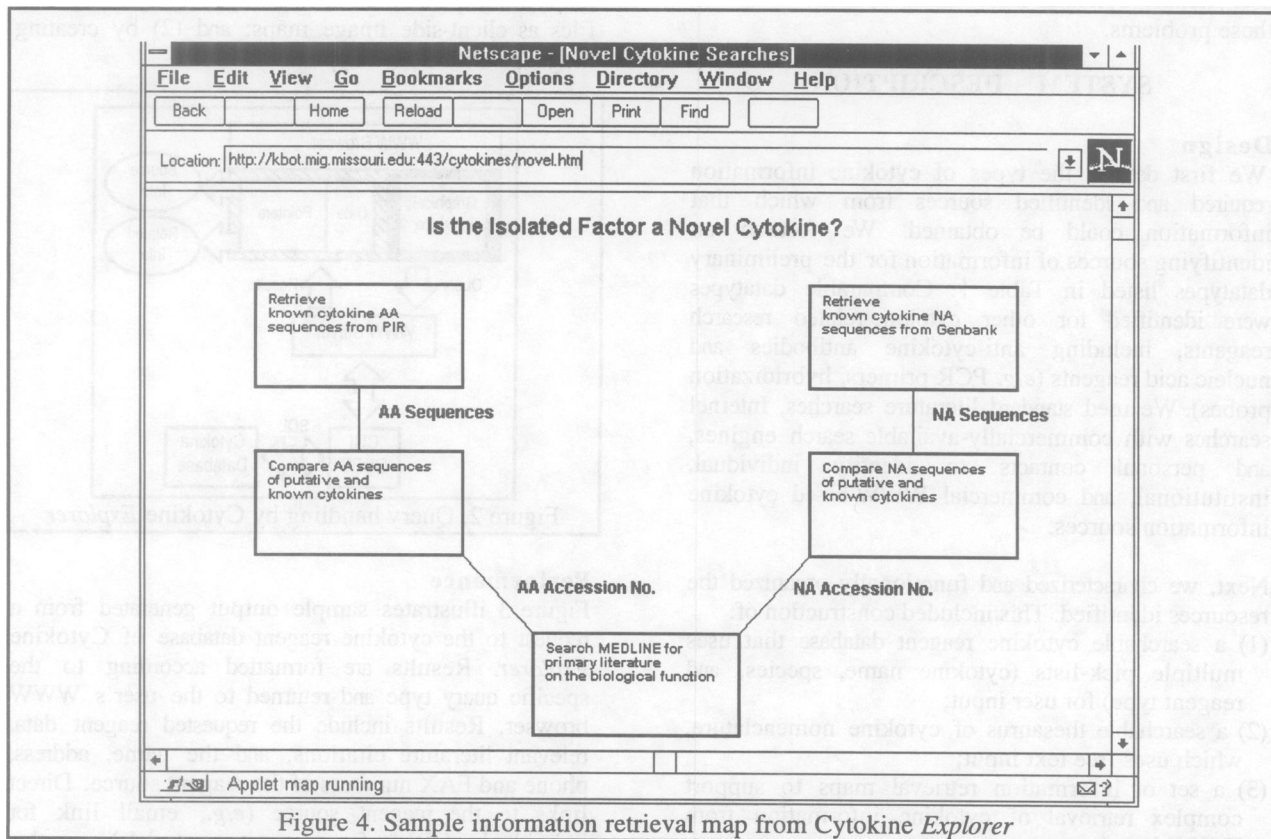


Figure 4. Sample information retrieval map from Cytokine Explorer

The WWW is a powerful means to accomplish these goals. Its near-universal accessibility and operating system- and hardware platform-independence make it an attractive vehicle for information retrieval applications for cytokine biology.

Cytokine *Explorer* has been designed to use these WWW attributes to provide a functional, flexible, and expandable solution to the salient issues of cytokine information retrieval. Various components of Cytokine *Explorer* effectively address core problem issues. The thesaurus addresses the problem of complicated cytokine nomenclature. The reagent database and links to other WWW-based sites address the problem of timely information retrieval. Many reagents from institutional sources, such as those from the University of Minnesota veterinary cytokine database, are not currently available commercially. Their inclusion, together with those from corporate sources, is important in our effort to provide for "one-stop shopping" for information on these research reagents.

Multiple knowledge engineering and implementation mechanisms (*e.g.*, mapping common requests for cytokine information onto datatypes and combinations of information resources) were used in both the reagent database and the information retrieval maps to provide integrated access to heterogeneous cytokine information sources. These system components incorporate the utility of multiple information resources which individually may be (1) narrowly focused with respect to content or (2) neither indexed nor organized in ways that obviously relate them to cytokine biology. It is our belief that further development of Cytokine *Explorer* will improve information access by research investigators and biomedical educators.

Acknowledgements

This work was supported in part by Grant No. 5 T15 LMO7089-04 from the National Library of Medicine and by the Veterinary Immunology Committee of the International Union of Immunological Societies. The authors thank K. Rugh for critical suggestions.

References

1. Barnett O, Shortliffe E, Chueh H, et al. Patient care applications on Internet. In: Ozbolt J, ed., *Proc of the 18th Ann SCAMC*; Washington, DC; Hanley & Belfus, Philadelphia 1994:1060.
2. Cimino JJ, Socratous SA, Grewal R: The informatics superhighway: prototyping on the World Wide Web. In Gardner RM, ed.: *Proc of the 19th Annual SCAMC*; New Orleans, LA; Hanley & Belfus, Philadelphia, 1995:111-115.
3. Protein Identification Resource. <http://ncbi.nlm.nih.gov/cgi-bin/pir>. 1995.
4. NCBI GenBank. <http://ncbi.nlm.nih.gov/cgi-bin/genbank>. 1995.
5. Oppenheim J, Rossio J, Gearing A: *Clinical Applications of Cytokines: Role in Pathogenesis, Diagnosis, and Therapy*. New York: Oxford University Press; 1993.
6. Ibelgauf H: *Dictionary of Cytokines*. New York, NY: Weinheim, 1995.
7. National Library of Medicine: Medical Subject Headings. Bethesda, MD, National Library of Medicine, 1994.
8. The *Journal of Immunology* Homepage. <http://199.170.0.125/JI>. 1996.
9. The Cell Press Homepage. <http://www.cell.com>. 1996.
10. The *Journal of Biological Chemistry* Online . http://highwire.stanford.edu/jbc/home/browse_search.html. 1996.
11. Casciari J, Fiege J, Sato H, Weinstein J: Reference database of cytokine structure and function. *Proc of the American Association for Cancer Research*, 1995;36:A470.
12. Cytokines Web. http://www.ocms.ox.ac.uk/~smb/cyt_web/. 1995.
13. Cytokine Homepage. <http://www.mb.uni-muenchen.de/groups/ibelgauf/cytokines.html>. 1995.
14. Cytokine Family cDNA Database: <http://cytokine.medic.kumamoto-u.ac.jp>. 1996.
15. Patrick T, Springer GK, Nicholas H, Ropelewski A: Information retrieval maps for biomedical research. In: Greenes R, Peterson H, Protti D, eds., *MEDINFO '95*, North-Holland, 1995;1443-1446.
16. Cytokine *Explorer*. <http://kbot.mig.missouri.edu:443/cytokines/explorer.html>. 1996.